

## 基因流介导的种间渐渗与物种界定\*

杜 芳, 徐 放

(北京林业大学计算生物学中心, 林木遗传育种国家工程实验室, 生物科学与技术学院, 北京 100083)

**摘要:** 物种界定是生物学中最基本的问题之一。近年来随着分子生物学技术的进步如分子条形码的发展, 物种界定也越来越引起人们的兴趣。界定一个物种或相似的一组物种时最重要的一个原则就是选择适合的分子标记。然而, 植物中广泛存在的不完全谱系筛选与种间渐渗却常常会阻碍准确鉴定物种。最近, 有关基因流介导的种间渗入和物种界定在理论和实验研究中都取得了重大进展。本文综述了基因流介导的物种形成; 评价了种间渐渗与不完全谱系筛选的区别; 最后总结出应该利用基因流速度较快的分子标记去鉴定物种。

**关键词:** 物种界定; 分子标记; DNA 条形码; 种间渐渗; 不完全谱系筛选; 基因流

中图分类号: Q 949, Q 943

文献标识码: A

文章编号: 2095-0845(2012)03-257-06

## Gene Flow Dependent Introgression and Species Delimitation

DU Fang, XU Fang

(Center for Computational Biology, National Engineering Laboratory for Tree Breeding, College of Bioscience and Technology, Beijing Forestry University, Beijing 100083, China)

**Abstract:** Species delimitation is one of the most fundamental issues in biology and has recently drawn significant interest. A main reason for the increasing interests was the barcoding initiative associated rapid development of molecular techniques. One of the most important principles to diagnose species or species groups is to choose appropriate markers. However, incomplete linkage sorting and introgression, which are widespread phenomena in plants, present major obstacles in species delimitation. Recently, significant progress in our understanding of gene flow dependent introgression and species delimitation has been made both theoretically and empirically. In this paper, we reviewed the gene flow mediated speciation; evaluated the difference of introgression and incomplete linkage sorting; and finally concluded that species delimitation should be more effective with markers experiencing high levels of gene flow.

**Key words:** Species delimitation; Molecular Marker; DNA barcoding; Species introgression; Incomplete linkage sorting; Gene flow

### Gene flow mediated speciation

Speciation refers to the evolution in reproductive barriers (as well as phenotypic, behavioral and genetic differences) between populations, eventually leading to distinct species (Coyne and Orr 1997; Rieseberg *et al.*, 2006 and references therein).

However, the mechanisms by which species are formed remain incompletely understood and the topic of intense research and debates. One major reason for the continued debates relates to different opinion on species concepts (Wiens, 2004).

Species concepts have played a major role in

\* Foundation item: Young Scientist Fund (2010BLX01) and the Fundamental Research Funds (YX2011-23) in BJFU Research Fund for the Doctoral Program of Higher Education of China (2011DD14120014) and an open funding (K1004) from Beijing Normal University to FKD

Received date: 2011-12-16, Accepted date: 2012-02-02

作者简介: 杜 芳 (1981-) 女, 讲师, 主要从事植物群体遗传学及数量遗传学研究, 特别关注森林物种间的基因流和种间渐渗等问题。E-mail: dufang@bjfu.edu.cn

evolutionary biology during the past 250 years, as summarized by De Queiroz (2007) in his review paper “Ernst Mayr and the modern concept of species”. The term “species” originally came from the Latin word for “kind” and its use has been made more precise following the work of Carolus Linnaeus (1753 and 1758). In this view, species is “the basic unit of biological classification” (Flexner and Hauck, 1993). A major shift took place after the seminal publication of Charles Darwin’s *On the Origin of Species* (1859). The modern evolutionary synthesis provided the foundation for systematics and evolutionary biology (Dobzhansky, 1937; Mayr, 1942). Ernst Mayr, the coiner of the biological species concept (BSC), proposed that species are ‘*groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups*’. This definition presents two characteristics of species: (i) that they have a continuous gene pool (i.e. all individuals can, or have the potential to, interbreed) and (ii), that the individuals of a species are reproductively isolated from individuals of other species (Mayr 1963; Niklas 1997). This definition suggests that interspecific gene flow should be low. Although the BSC is widely accepted by zoologists, lots of botanists prefer using morphological characters as main features to diagnose species and criticized the use of reproductive barrier for species delineation because such a barrier is not as strict in plants as in animals (e.g. many plant taxa are potentially interfertile or parthenogenesis is common in plants). In addition, Mayr pointed out in further statements later that ‘*the steady and high genetic input caused by gene flow is the main factor responsible for genetic cohesion among the populations of a species*’ (Mayr, 1963). This later argument emphasizes the importance of intraspecific gene flow, which unites all individuals and populations of one species together. While most researchers agreed that interspecific gene flow ought to be limited to keep species distinct, arguments were made against the importance of intraspecific gene flow, as species

cohesion did not seem to be always dependent on intraspecific gene flow (Ehrlich and Raven, 1969).

Recently Morjan and Rieseberg (2004) demonstrated that intraspecific gene flow, even when limited, is essential to keep species genetically coherent, while acknowledging that other evolutionary forces such as selection are also important in maintaining species cohesion. Both components, inter- and intraspecific gene flow, therefore lie at the root of the biological species concept (Mayden, 1997). However, their interaction and the way they affect species delimitation have drawn attention only very recently (Petit and Excoffier 2009; Du *et al.*, 2011)

Species delimitation has involved many methods to identify the actual boundaries of species and at the same time determine the number of species (De Queiroz, 2007). Delimiting species has traditionally relied on morphological characters supplemented with geographic and ecological information (Briggs and Walters, 1997). Interest for species delimitation has fluctuated through time: it drew a lot of attention in the middle of the last century, thanks to the emergence of modern systematics (Sites and Marshall, 2004), and now experiences a phase of renaissance thanks to the rapid development of molecular technologies (Wiens, 2004). In particular, DNA barcoding (the use of a short standardized DNA sequence to identify and discover species) has attracted much attention (Hebert *et al.*, 2003, 2004; Hollingsworth *et al.*, 2011; Li *et al.*, 2011). A key point for DNA barcoding and more generally for species delimitation is to choose a “good” marker showing species-specific variation.

### Incomplete lineage sorting and introgression

However, numerous studies have revealed shared DNA polymorphisms between closely related species. This situation can be caused by two main reasons: (1) retention of ancestral polymorphisms, caused by incomplete lineage sorting (also called sharing of ancestral polymorphisms) during and following speciation (Heckman *et al.*, 2007; Willyard

*et al.*, 2009 and references therein); (2) introgression, caused by genetic exchange after secondary contact between two previously geographically separated species (Liston *et al.*, 1999; Gay *et al.*, 2007). Distinguishing these two mechanisms is difficult; the most common approach is to use coalescent modeling to compare divergence time and ancestral population sizes of the two species. Several studies have used this approach (Hey, 2001; Rannala and Yang, 2003; Smith and Farrell, 2005; Burgess and Yang, 2008; Joly *et al.*, 2009). However, for closely related species that have diverged very recently, this type of approach seems to have limited utility, unless spatial information is taken into account (McGuire *et al.*, 2007). Incomplete lineage sorting and introgression can be differentiated by studying the geographic variation and demographic history of the species using molecular markers. If shared polymorphisms are randomly distributed, then retention of ancestral polymorphisms might be involved. However, if shared haplotypes occur only in sympatric populations, then introgression is more likely (Fig. 1).

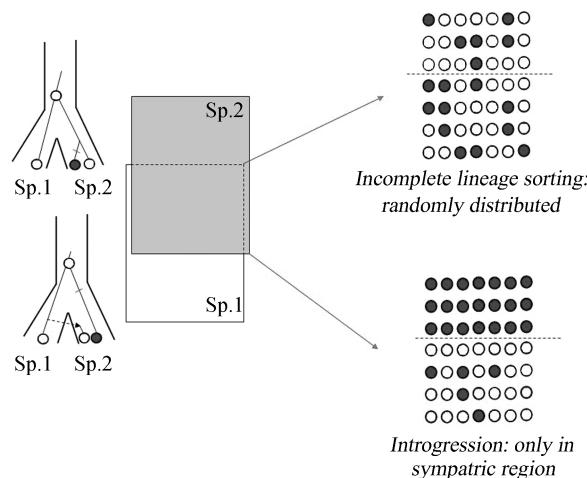


Fig. 1 Possible interpretations of the mechanisms underlying cases of shared haplotypes between species: If shared haplotypes are randomly distributed, incomplete lineage sorting is more likely (top left). If introgression is involved, following contact between species, shared haplotypes will be restricted to the areas of sympatry (bottom left). In the illustration, only the distribution of haplotypes of species 2 is illustrated

Introgression is a widespread phenomenon with potentially profound evolutionary consequences (Anderson and Hubricht, 1938; Anderson, 1953; Anderson and Stebbins, 1954; Rieseberg and Brunsfeld, 1992). Edgar Anderson defined introgression as the process of infiltration of genes from one species to another through regular mating events involving backcrosses with one of the parental species. The introgression process can be divided into three steps: generation of F1 hybrids; backcrossing with one or both of the parents; incorporation of this new genetic variation into the genome of the backcrossing species, possibly following screening by natural selection (Anderson and Hubricht, 1938; Anderson, 1953; Anderson and Stebbins, 1954). Indeed, individuals with introgressed genetic materials can selectively retain (or “filter”) advantageous genes, while disadvantageous genes can be eliminated by purifying selection (Key, 1968; Harrison, 1986). The possibilities offered by introgression were realized early on by breeders willing to incorporate in a domesticated species a given attribute of a wild relative (Bessey, 1906; Gur and Zamir, 2004). Introgression processes may differ for different genomes. In the nuclear genome, the F1 hybrid gets 50% of the genes from each parent, and the proportion of additional introgressed genetic material is halved after every generation of backcrossing. For uniparentally inherited genomes, the situation is strikingly different. Each F1 hybrid receives a complete unaltered version of the genome from one of its parent, so that there is no dilution of the contribution of the donor species or population after several generations of backcrossing (Fig. 2). The formation of introgression can be very quick (several generations), in contrast, incomplete lineage sorting typically represents much more ancient events.

Note that genetic drift is greatly reduced in a subdivided population compared to a single random mating population of similar census size (Wright, 1943; Gilpin, 1991). Hence, under the condition of similar population size reduced intraspecific gene flow,

which implies increased genetic drift and hence increased subdivision, means that sorting of ancestral variation will take longer, making species diagnosis

more challenging. Thus, species delimitation should be easier if it were based on molecular markers experiencing high rates of gene flow (Fig. 3).

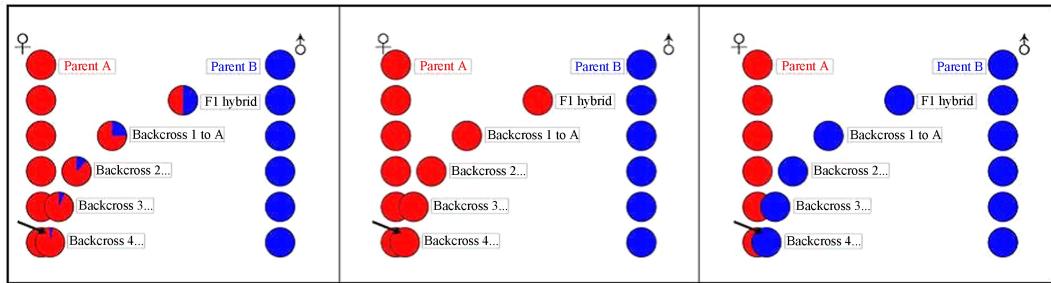


Fig. 2 Illustration of the introgression process for three different genomes: Biparentally inherited nuclear genome (left); maternally inherited organelle genome (middle) and paternally inherited organelle genome (right). The red circles represent genes with maternal ancestry and the blue circles represent genes with paternal ancestry in the offspring. For the nuclear genome (left) the genetic material inherited from the donor parent is reduced to 1/32 after four generations of backcrossing. For the maternally inherited genomes, the genetic material from the father does not contribute at all (middle). In contrast, for a paternally inherited genome, the offspring retains the entire genome from the father, there is no dilution effect (right)

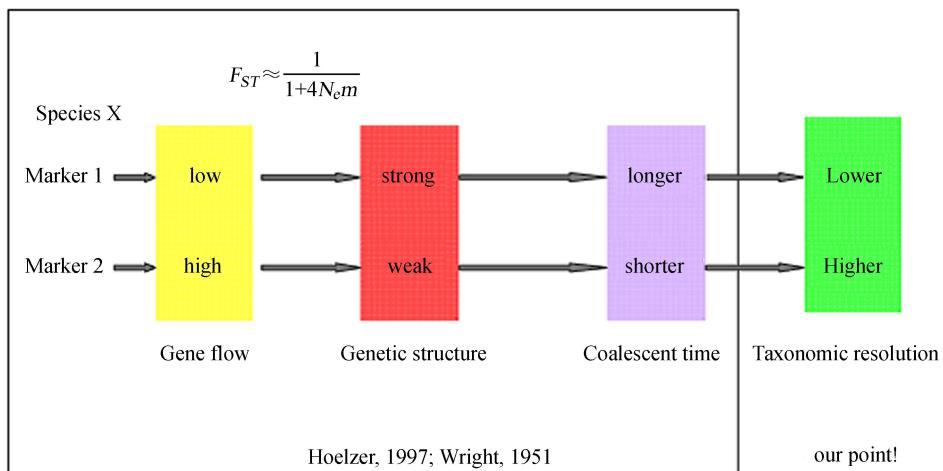


Fig. 3 Species delimitation ability at two markers experiencing contrasted rates of gene flow under the condition of similar population size. The marker with lower levels of gene flow (marker 1) has lower taxonomic resolution than the marker with higher gene flow (marker 2)

## Recent progress on gene flow dependant introgression and species delimitation

To date, many empirical studies have already demonstrated that incomplete linkage sorting is widespread in plants (Du *et al.*, 2009; Zhou *et al.*, 2010; Wang *et al.*, 2011a; Palma-Silva *et al.*, 2011), however, studies on introgression in plant is lacking (but see Arnold *et al.*, 2010). Recently, significant progress in our understanding of introgression has been made with the development of a neu-

tral demo-genetic model (Currat *et al.*, 2008). This model predicts that, when one species invades an area already occupied by a related species, introgression of neutral genes takes place mainly from the native species towards the invading species. In addition, following contacts between two hybridizing species, the model predicts that introgression should be particularly frequent for genome components experiencing little gene flow. In line with this neutral model, Petit and Excoffier (2009) suggested that

markers experiencing high rates of gene flow should be better suited for species delimitation than those experiencing low rates of gene flow, in part because high rates of intra-specific gene flow can prevent introgression.

Empirical studies also support the above predictions in both gymnosperm and angiosperm plants. Du *et al.* (2011) use molecular markers from two organelle genomes ( mtDNA and cpDNA ) with contrasting rates of gene flow to examine genetic exchanges between two morphologically distinct spruce *Picea* species growing in the Qinghai-Tibetan Plateau. They found that all sympatric populations of the expanding species had received their maternally inherited mitochondrial DNA ( mtDNA ) ( transferred by seed, low gene flow ) from the resident species, whereas for paternally inherited chloroplast ( cpDNA ) ( transferred by pollen and seeds, high gene flow ) introgression is more limited and not strictly unidirectional ( See their schematic model in Fig. 1 of Du *et al.*, 2011 ). In angiosperm plants, however, after comparative analysis of a large dataset on both chloroplast DNA (*rbcL*, *matK* and *trnH-psbA*) and nuclear internal transcribed spacer (ITS), The China Plant BOL Group *et al.* (2011) discovered that the later performed relatively well in angiosperm plant species delimitation. This conclusion based on a large dataset represents another step forward towards routine use of DNA barcoding ( Hollingsworth *et al.*, 2011 ) as well as using markers with fast rate of gene flow to diagnose species ( Wang *et al.*, 2011b ).

## Conclusion

Recent progress on both theoretical and empirical studies suggested that the important role of gene flow should not be ignored no matter the study focus is on the species demographic history or diagnostics. If the studies were focusing on revealing the phylogeographic history of the species then the markers with low rate of gene flow should be used, i. e. mtDNA in gymnosperm or organic ( cp ) DNA in angiosperm. However, if the studies were designed to

delimitate related species or species groups ( barcoding for example ), then the markers with fast rate of gene flow should be chosen i. e. cpDNA in gymnosperm or nuclear DNA in angiosperms.

**Acknowledgements:** This review was part of FK Du's doctoral thesis. The authors thank Prof. XueJue Ge, South China Botanical Garden; Prof. Jianquan Liu in Lanzhou University; Prof Zhixiang Zhang from Beijing Forestry University; Martin Lascoux from Uppsala University, Sweden and Rémy J. Petit from INRA, France who give comments for this manuscripts.

## References:

- Anderson E, Hubricht L, 1938. Hybridization in tradescantia. III. The evidence for introgressive hybridization [ J ]. *American Journal of Botany*, **25** ( 6 ): 396—402
- Anderson E, 1953. Introgressive hybridization [ J ]. *Biological Reviews*, **28** ( 3 ): 280—307
- Anderson E, Stebbins GLJ, 1954. Hybridization as an evolutionary stimulus [ J ]. *Evolution*, **8** ( 4 ): 378—388
- Arnold ML, Tang SX, Knapp SJ *et al.*, 2010. Asymmetric introgressive hybridization among louisiana iris species [ J ]. *Genes*, **1**: 9—22
- Bessey CE, 1906. Crop improvement by utilizing wild species [ J ]. *Journal of Heredity*, os-2 ( 1 ): 112—118
- Briggs D, Walters SM, 1997. *Plant Variation and Evolution* [ M ]. Cambridge: Cambridge University Press
- Burgess R, Yang ZH, 2008. Estimation of hominoid ancestral population sizes under bayesian coalescent models incorporating mutation rate variation and sequencing errors [ J ]. *Molecular Biology Evolution*, **25** ( 9 ): 1979—1994
- China Plant BOL Group, Li DZ, Gao LM *et al.*, 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants [ J ]. *Proceedings of the National Academy of Science*, **108** ( 49 ): 19641—19646
- Coyne JA, Orr HA, 1997. “Patterns of speciation in *Drosophila*” revisited [ J ]. *Evolution*, **51** ( 1 ): 295—303
- Currat M, Ruedi M, Petit RJ *et al.*, 2008. The hidden side of invasions: massive introgression by local genes [ J ]. *Evolution*, **62** ( 8 ): 1908—1920
- De Queiroz K, 2007. Species concepts and species delimitation [ J ]. *Systematic Biology*, **56** ( 6 ): 879—886
- Dobzhansky T, 1937. *Genetics and the Origin of Species* [ M ]. New York: Columbia University Press, 364
- Du FK, Petit RJ, Liu JQ, 2009. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers [ J ]. *Mo-*

- lecular Ecology*, **18** (7) : 1396—1407
- Du FK, Peng XL, Liu JQ et al., 2011. Direction and extent of organellar DNA introgression between two spruce species in the Qinghai-Tibetan Plateau [J]. *New Phytologist*, **192** (4) : 1024—1033
- Ehrlich PR, Raven P, 1969. Differentiation of populations [J]. *Science*, **165** (3899) : 1228—1232
- Flexner SB, Hauck LC, 1993. *Random House Webster's Unabridged Dictionary* [M]. German: Random House
- Gay L, Neubauer G, Zagalska-Neubauer M et al., 2007. Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact [J]. *Molecular Ecology*, **16** (15) : 3215—3227
- Gilpin M, 1991. The genetic effective size of a metapopulation [J]. *Biological Journal of the Linnean Society*, **42** (1—2) : 165—175
- Gur A, Zamir D, 2004. Unused natural variation can lift yield barriers in plant breeding [J]. *Plos Biology*, **2** (10) : e245
- Harrison RG, 1986. Pattern and process in a narrow hybrid zone [J]. *Heredity*, **56** (3) : 337—349
- Hebert PDN, Ratnasingham S, deWaard JR, 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species [J]. *Proceedings of the Royal Society B: Biological Sciences*, **270** (Suppl 1) : S96—S99
- Hebert PDN, Stoeckle MY, Zemlak TS et al., 2004. Identification of birds through DNA barcodes [J]. *Plos Biology*, **2** (10) : e312
- Heckman KL, Mariani CL, Rasolofoson R et al., 2007. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*) [J]. *Molecular Phylogenetics and Evolution*, **43** (2) : 353—367
- Hey J, 2001. The mind of the species problem [J]. *Trends in Ecology & Evolution*, **16** (7) : 326—329
- Hollingsworth PM, Graham SW, Little DP, 2011. Choosing and using a plant DNA barcode [J]. *Plos One*, **6** (5) : e19254. doi: 10.1371/journal.pone.0019254
- Joly S, McLenaghan PA, Lockhart PJ, 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting [J]. *American Naturalist*, **174** (2) : E54—E70
- Key KHL, 1968. The concept of stasipatric speciation [J]. *Systematic Zoology*, **17** (1) : 14—22
- Liston A, Robinson WA, Piñero D et al., 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences [J]. *Molecular Phylogenetics and Evolution*, **11** (1) : 95—109
- Mayden RL, 1997. “A hierarchy of species concepts: the denouement in the saga of the species problem” [J]. *Species the Units of Biodiversity*, **54** : 381—424
- Mayr E, 1942. *Systematics and the Origin of Species* [M]. New York: Columbia University Press
- Mayr E, 1963. *Animal Species and Evolution* [M]. Cambridge MA: Harvard University Press
- McGuire JA, Linkem CW, Koo MS et al., 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards [J]. *Evolution*, **61** (12) : 2879—2897
- Morjan CL, Rieseberg LH, 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles [J]. *Molecular Ecology*, **13** (6) : 1341—1356
- Niklas KJ, 1997. *The Evolutionary Biology of Plants* [M]. Chicago: The University of Chicago Press
- Palma-silva C, Wendt T, Pinheiro F et al., 2011. Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs [J]. *Molecular Ecology*, **20** (15) : 3185—3201
- Petit RJ, Excoffier L, 2009. Gene flow and species delimitation [J]. *Trends in Ecology & Evolution*, **24** (7) : 386—393
- Rannala B, Yang ZH, 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci [J]. *Genetics*, **164** (4) : 1645—1656
- Rieseberg LH, Brunsfeld SJ, 1992. Molecular evidence and plant introgression [A]. In: Soltis DE, Soltis PS, Doyle JJ (eds). *Molecular Systematics of Plants* [M]. New York: Chapman and Hall Inc., 151—176
- Rieseberg LH, Wood TE, Baack EJ, 2006. The nature of plant species [J]. *Nature*, **440** (7083) : 524—527
- Sites JW, Marshall JC, 2004. Operational criteria for delimiting species [J]. *Annual Review of Ecology, Evolution, and Systematics*, **35** (1) : 199—227
- Smith CI, Farrell BD, 2005. Phylogeography of the longhorn cactus beetle *Moneilema appressum* LeCont (Coleoptera: Cerambycidae) was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? [J]. *Molecular Ecology*, **14** (10) : 3049—3065
- Wang J, Wu YX, Ren GP et al., 2011a. Genetic Differentiation and Delimitation between Ecologically Diverged *Populus euphratica* and *P. pruinosa* [J]. *PLoS One*, **6** (10) : e26530. doi:10.1371/journal.pone.0026530
- Wang Q, Yu QS, Liu JQ, 2011b. Are nuclear loci ideal for barcoding plants? A case study of genetic delimitation of two sister species using multiple loci and multiple intraspecific individuals [J]. *Journal of Systematics and Evolution*, **49** (3) : 182—188
- Wiens JJ, 2004. What is speciation and how should we study it? [J]. *The American Naturalist*, **163** (6) : 914—923
- Wilyard A, Cronn R, Liston A, 2009. Reticulate evolution and incomplete lineage sorting among the *Ponderosa pines* [J]. *Molecular Phylogenetics and Evolution*, **52** : 498—511
- Wright S, 1943. Isolation by distance [J]. *Genetics*, **28** (2) : 114—138
- Zhou YF, Abbott RJ, Jiang ZY et al., 2010. Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China [J]. *Evolution*, **64** (8) : 2342—2352